

QUALIDADE E EQUIDADE NO ENSINO FUNDAMENTAL BRASILEIRO*

Ângela Albernaz

Do Departamento de Economia da PUC-Rio

Francisco H. G. Ferreira

Do Departamento de Economia da PUC-Rio

Creso Franco

Do Departamento de Educação da PUC-Rio

Este artigo estima uma função de produção educacional para o ensino fundamental brasileiro, com base nos dados do Sistema Nacional de Avaliação da Educação Básica (Saeb) para a 8ª série em 1999. Utilizando modelos hierárquicos lineares, de forma a evitar vieses de agregação, a estimação sugere que cerca de 80% da variância de desempenho médio entre as escolas devem-se a diferenças na composição socioeconômica de seus alunos. Como em outros estudos internacionais, a rede a que a escola pertence está relacionada de forma importante com o desempenho de seus alunos. Não obstante, e em contraste com a maior parte da literatura internacional, outras variáveis escolares, tais como a escolaridade dos professores e a qualidade da infra-estrutura física, também afetam o desempenho dos estudantes. Alguns fatores positivos para a eficácia média das escolas reduzem a sua “equidade”.

1 INTRODUÇÃO

O sistema educacional de um país é um conjunto de instituições fundamental para o seu desenvolvimento socioeconômico. Desde Arrow (1962) e Uzawa (1965) até Lucas (1988) e Becker, Murphy e Tamura (1990), a acumulação de capital humano, em geral, e a educação, em particular, ocupam posição central em nossa visão sobre o processo de crescimento econômico. A educação também exerce um papel crucial na determinação da distribuição dos frutos desse crescimento, como sabemos a partir de Mincer (1958), passando por Tinbergen (1975). No Brasil, particularmente, a importância da educação para explicar o alto nível de desigualdade vem sendo apontada desde Langoni (1973).¹

Há, portanto, grande interesse acadêmico e de política pública no sentido de uma melhor compreensão dos determinantes do desempenho educacional dos alunos brasileiros. Em particular, interessa quantificar os efeitos dos vários insumos educacionais — tais como a escolaridade e a experiência dos professores, e os vários aspectos da infra-estrutura escolar — sobre o aprendizado. O objetivo deste trabalho é contribuir para essa maior compreensão, através da estimação de uma função de produção educacional para o ensino fundamental brasileiro.

* Os autores agradecem ao professor Naércio Aquino Menezes-Filho e a dois pareceristas anônimos por suas excelentes sugestões. É claro que continuamos sendo os únicos responsáveis pelos erros que porventura permaneçam no artigo.

1. Para mais detalhes, ver Barros, Henriques e Mendonça (2000).

A “função de produção educacional” é o nome dado pelos economistas à relação existente entre uma série de “insumos” ao processo educacional e o seu “produto”. Desde a publicação do Relatório Coleman sobre a qualidade das escolas públicas e privadas nos Estados Unidos [Coleman *et alii* (1966)], desenvolveu-se uma literatura considerável sobre o tema, tanto em economia como na área da educação. A forma geral da função de produção educacional é dada pela equação (1):

$$y = F(c, m, g, p, s) \quad (1)$$

onde y denota o desempenho dos alunos, que pode depender de uma série de fatores, agrupados em cinco categorias: características pessoais do aluno (c), tais como raça e gênero; características de suas famílias (m), tais como renda ou outra medida de seu nível socioeconômico; características de seus colegas na escola (g); características dos seus professores (p), tais como escolaridade, salário ou experiência; e outras características escolares (s).²

No caso brasileiro, a análise econômica desta relação entre insumo e produto do processo educacional sofria, tradicionalmente, de um sério problema, referente à mensuração da variável produto (y): até a publicação dos primeiros resultados dos exames padronizados do Sistema Nacional de Avaliação da Educação Básica (Saeb), em 1997, a única variável disponível para medir desempenho educacional era o número de séries completadas pelo aluno (anos de escolaridade).³ Ainda que estudos com base nesta variável tenham contribuído para uma compreensão inicial da problemática educacional brasileira [ver Barros, Mendonça e Santos (1999)], esses mesmos estudos reconheceram as severas limitações da variável “anos de escolaridade” como medida de aprendizado.⁴

A disponibilização dos microdados do Saeb⁵ tornou possível, pela primeira vez no Brasil, a investigação dos determinantes de uma medida de desempenho escolar com base em rendimentos de alunos em testes padronizados de conhecimento. Além de incluir informações sobre o rendimento dos alunos nesses testes, a base de dados do Saeb abrange ainda informação sobre todas as outras categorias de determinantes incluídos na equação (1), constituindo, assim, um recurso potencialmente muito valioso para o estudo da eficácia, eficiência e equidade da

2. Para uma resenha clássica desta literatura, ver Hanushek (1986).

3. Vale mencionar que diversos estudos utilizaram outras medidas elaboradas a partir dessa variável, como taxas de fluxo ou defasagem idade-série.

4. O principal problema com essa variável é que ela não inclui qualquer informação sobre a qualidade de cada ano de estudo, nesse contexto entendido como o domínio de conteúdo, por parte dos alunos.

5. Assim como de seu equivalente para o ensino superior (o “Provão”) e o Enem.

escola brasileira. Estudos na área de educação como os de Fletcher (1997), Barbosa e Fernandes (2001) e Franco, Mandarino e Ortigão (2001) têm confirmado a importância dessa base de dados.

Neste artigo, estimamos uma função de produção educacional para o Brasil, aplicando modelos hierárquicos lineares aos microdados do Saeb de 1999. Ao contrário de qualquer estudo anterior de que temos conhecimento, investigamos a contribuição de diferentes variáveis escolares (e de professores) tanto sobre uma medida de eficácia, quanto para uma medida de equidade da prática educacional brasileira.

Dentre os resultados encontrados, destacamos três. Primeiro, como em outros países, a variância do desempenho entre as escolas brasileiras deve-se principalmente a diferenças no nível socioeconômico médio dos alunos das escolas, refletindo um importante efeito de seleção da clientela. Segundo, uma vez controlado esse efeito, diferenças na quantidade e qualidade dos insumos escolares ainda respondem por uma parcela significativa da diferença de desempenho entre as escolas. Ao contrário de resultados encontrados para vários outros países, tanto a qualidade dos professores quanto a qualidade da infra-estrutura física das escolas afetam o rendimento de forma significativa.⁶ Terceiro, mesmo controlando por todos esses fatores, no Brasil o desempenho médio da escola particular supera o da escola pública.

O trabalho está organizado em quatro seções além desta introdução. Na Seção 2, apresentamos os dados e a metodologia utilizada. Na Seção 3, discutem-se os resultados para o modelo estimado sem as características escolares. A Seção 4 apresenta o modelo completo e a Seção 5 apresenta nossas conclusões.

2 DADOS E METODOLOGIA

Este artigo utiliza a base de dados do Saeb de 1999 — uma amostra probabilística de alunos brasileiros. Em particular, utilizamos a amostra da 8ª série, que inclui informações sobre o desempenho em ciências, geografia, história, português e matemática de 89.671 alunos pertencentes a 2.588 escolas. O tamanho médio da amostra é de aproximadamente 35 alunos por escola. O plano amostral do Saeb envolve estratificação e conglomerados. Na 8ª série, as variáveis de estratificação foram: unidades da federação, dependência administrativa das escolas (estadual, municipal e particular) e tamanho da escola. Os conglomerados são representa-

6. A pouca relevância das variáveis escolares tem sido uma característica polêmica, mas persistente, da literatura internacional sobre funções de produção de educação. De acordo com a Tabela 3.23.1, p. 303 do Relatório Coleman, menos de 2% da variância total do desempenho dos alunos (brancos e negros) são atribuídos a características escolares, enquanto menos de 4% são atribuídos a características dos professores. Em uma resenha mais recente, Hanushek (1989) apresenta um resumo das estimativas dos coeficientes dos gastos escolares sobre o desempenho dos alunos em 187 estudos, chegando à conclusão de que a única variável cuja relevância para o aprendizado dos alunos parece ser realmente robusta é a experiência do professor.

dos pelas escolas sorteadas para compor a amostra. Dentro das escolas, a seleção de turmas foi realizada utilizando-se uma amostragem aleatória simples. Cada aluno foi testado em apenas uma disciplina.⁷

Esse sistema de avaliação oferece uma série de informações que vai desde o desempenho dos alunos, medido pela nota obtida no teste, até as características dos próprios alunos, dos seus professores e das escolas onde eles estudam. A disponibilidade desses resultados nos estimula a avaliar quais as características dos alunos, de seus professores e das escolas onde eles estudam que favorecem um melhor desempenho deles.

A Tabela 1 apresenta as médias e os desvios-padrão das variáveis de rendimento escolar. Os erros-padrão reportados levam em consideração o desenho amostral com conglomerados e estratificação. A média de matemática foi escolhida como a média-base para todas as matérias. Portanto, os valores reportados para as demais disciplinas representam a média das variações dessas disciplinas em relação à média de matemática.

Como qualquer outra base de dados, o Saeb não é perfeito. Um de seus problemas, do ponto de vista da análise econômica, é a ausência de uma variável de renda familiar capaz de discriminar as diferenças de renda entre as famílias das crianças que estudam na mesma escola. Para contornar o problema, adotamos neste trabalho uma prática comum na análise educacional, que é a elaboração de uma variável *proxy* do nível socioeconômico do aluno, através da extração do componente principal de nove variáveis derivadas de respostas sobre educação (paterna e materna) e posse de bens e serviços do questionário familiar do Saeb. A variável de nível socioeconômico utilizada neste trabalho está altamente associada à medida de nível socioeconômico da Associação Brasileira dos Institutos de

TABELA 1
ESTATÍSTICAS DESCRITIVAS PARA AS PROVAS REALIZADAS PELOS ALUNOS DA 8ª SÉRIE

	Variável	Média	Desvio-padrão
Desempenho em matemática	Y_{ij}	254,72	51,40
Desvio médio de ciências em relação a matemática	$d_{cie_{ij}}$	-2,34	0,92
Desvio médio de geografia em relação a matemática	$d_{geo_{ij}}$	3,68	0,97
Desvio médio de história em relação a matemática	$d_{hit_{ij}}$	3,67	0,85
Desvio médio de português em relação a matemática	$d_{port_{ij}}$	-13,50	0,93

7. Alunos matriculados em escolas federais foram excluídos da população de referência do Saeb. Informações adicionais sobre o plano amostral do Saeb de 1999 podem ser obtidas em Bussab, Silva e Andrade (2000).

Pesquisa de Mercado (Abipeme) ($r = 0,88$), que foi validada a partir de estudos nos quais tanto o acesso à educação quanto aos bens e serviços são regredidos contra a renda dos informantes [Mattar (1995)]. A principal diferença entre a variável de nível socioeconômico que utilizamos e a variável Abipeme consiste na gradação mais detalhada da variável que utilizamos. Esta característica garante a sensibilidade da variável para captar diferenças relativamente pequenas no nível socioeconômico dos alunos, indispensável nas investigações que envolvem a estimação de regressões a partir dos dados de cada escola.⁸

Um segundo problema com os dados do Saeb é que seu desenho não permite a criação de uma variável de *diferencial* dos testes de rendimento (por exemplo, da 7ª para a 8ª série). Tal estimativa, chamada de medida de valor agregado da educação, é reconhecida teoricamente como a mais apropriada, pois corresponde ao aprendizado do aluno no período de tempo considerado [Hanushek (1989)]. Na sua ausência, utilizamos a melhor variável disponível, que é o *nível* do teste de rendimento do aluno.

A Tabela 2 mostra a definição e a descrição das variáveis explicativas utilizadas na modelagem, apresentada na próxima seção.

Apresentaremos a seguir a metodologia econométrica a ser utilizada na estimação da equação (1), com base nos dados do Saeb, que é a elaboração de um modelo hierárquico linear. Os modelos hierárquicos lineares — também conhecidos como modelos multinível — são modelos de coeficientes aleatórios, apropriados para a análise de variáveis aleatórias cujas distribuições sobre a população não são independentes e idênticas, mas sim agrupadas de forma não-aleatória em subgrupos relevantes. Dados educacionais são um exemplo clássico, já que se compõem de variáveis descrevendo os alunos, agrupados em unidades maiores — as escolas — onde eles interagem. Assim sendo, é demasiadamente restritivo impor que os coeficientes relacionando as variáveis independentes à variável dependente sejam constantes entre os subgrupos (no caso educacional, entre as escolas), ou que os resíduos sejam iid.

Os modelos hierárquicos lineares combinam a flexibilidade da estimação de equações separadas para cada escola com o fato de a variação desses coeficientes

8. Em princípio, a análise do componente principal pressupõe medidas contínuas, já que se viabiliza a partir da matriz de correlação com base em coeficientes de correlação de Pearson. Por outro lado, as variáveis disponíveis são ordinais e contagens. A despeito disso, justifica-se conceitualmente o bom funcionamento da variável elaborada — empiricamente já justificado pela alta correlação com a variável Abipeme — considerando-se que a estimação da matriz de correlação, embora tenda a subestimar a efetiva correlação entre os pares de variáveis, captura a estrutura subjacente aos dados, na medida em que os coeficientes de correlação agrupam as variáveis mais correlacionadas entre si. Por isso, não é de se estranhar que diversos autores façam uso da análise de componentes principais com variáveis ordinais. Fica para uma agenda de pesquisa futura o uso de abordagens que equacionem, simultaneamente, a análise da dimensionalidade dos dados gerados por questionários e a estimação de fator latente a partir da matriz de correlação estimada por métodos não-paramétricos [ver Molenaar (1997)].

TABELA 2
DEFINIÇÃO E ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS EXPLICATIVAS

Variáveis	Descrição	Média	Desvio-padrão	Mínimo	Máximo
Aluno					
Repete_1	1 = repetiu uma vez e 0 = caso contrário	0,24	0,43	0,00	1,00
Repete_2	1 = repetiu duas vezes e 0 = caso contrário	0,13	0,34	0,00	1,00
Repete_3	1 = repetiu três vezes e 0 = caso contrário	0,05	0,22	0,00	1,00
Repete_4	1 = repetiu quatro vezes e 0 = caso contrário	0,02	0,14	0,00	1,00
Pardo	1 = pardo ou mulato e 0 = caso contrário	0,37	0,48	0,00	1,00
Preto	1 = preto e 0 = caso contrário	0,06	0,24	0,00	1,00
Masc	1 = sexo masculino e 0 = sexo feminino	0,46	0,50	0,00	1,00
Nse	Nível socioeconômico do aluno	0,00	1,12	-2,84	9,20
Escola					
Particular	1 = rede particular e 0 = rede pública	0,25	0,43	0,00	1,00
Sala	1 = salas de aula arejadas e 0 = salas de aula não arejadas	0,88	0,32	0,00	1,00
Ruído	1 = ruído prejudica e 0 = ruído não prejudica	0,30	0,46	0,00	1,00
Recursos	1 = insuficiência de recursos financeiros e 0 = não há insuficiência de recursos financeiros	0,71	0,34	0,00	1,00
Faltas	1 = falta de aula, por parte dos alunos e 0 = não há falta de aula, por parte dos alunos	0,32	0,33	0,00	1,00
Escprof	Nível de escolaridade médio dos professores da escola. Agregado da resposta dos professores à pergunta sobre seu nível de escolaridade. Valores extremos: 3 = concluiu o ensino superior e 0 = não concluiu o ensino fundamental	2,83	0,29	1,50	3,00
Salário	Variável ordinal correspondendo a dez faixas salariais. Valores extremos: 0 = até 1 sm e 9 = mais de 20 sm — 9 = professores da escola que recebem mais de R\$ 2.720	4,14	1,59	0,00	9,00
Prop_rep	Proporção de alunos repetentes	0,44	0,24	0,00	1,00
Nsemédio	Nível socioeconômico médio da escola	0,02	0,77	-2,11	2,98

Obs.: A variável nsemédio foi obtida por agregação, para a base da escola, da média da variável nse. As variáveis sala, ruído, faltas, escprof e salário foram obtidas por agregação, para a base da escola, da média das medidas existentes na base do professor.

entre as escolas poder ser estimada em um segundo nível do modelo. O modelo hierárquico geral de dois níveis⁹ é representado pelas seguintes equações:

Nível 1

$$Y_{ij} = \beta_{0j} + \sum_{q=1}^Q \beta_{qj} X_{qij} + r_{ij} \quad r_{ij} \sim N(0, \sigma^2)$$

Nível 2

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj} \quad u_{qj} \sim N(0, \tau_{qq'})$$

onde i indexa os alunos e j indexa as escolas. Existem Q coeficientes (potencialmente) aleatórios no nível 1, indexados por q . Cada um deles determinado por S_q variáveis independentes, indexadas por s . O efeito dessas variáveis sobre cada coeficiente é estimado no nível 2. Para qualquer par de coeficientes aleatórios q e q' , $\text{cov}(u_{qj}, u_{q'j}) = \tau_{qq'}$. Os erros de nível 1 e de nível 2 são supostos independentes: $\text{cov}(r_{ij}, u_{qj}) = 0, \forall_{q,i,j}$. O algoritmo de estimação do modelo baseia-se na maximização da função de verossimilhança conjunta dos dois níveis. As referências estatísticas clássicas para esses modelos são Bryk e Raudenbush (1992) e Goldstein (1995).

No caso específico da estimação de funções de produção educacional, a aplicação desses modelos, além de estatisticamente mais recomendável, tem efeitos potencialmente sérios sobre uma das conclusões mais importantes da literatura: a pouca importância relativa das variáveis educacionais *vis-à-vis* as variáveis familiares na determinação do desempenho escolar. Diversos estudos que utilizam mínimos quadrados ordinários (MQO) apontam a existência de uma forte relação entre as características dos alunos e de suas famílias e o desempenho escolar [Summers e Wolfe (1977), Hanushek (1989), Mizala e Romaguera (1998), Mizala, Romaguera e Reinaga (1999) e Barros, Mendonça e Santos (1999)], mas concluem que não há uma relação robusta entre as variáveis escolares e as características dos professores, por um lado, e o rendimento dos alunos, por outro [Hanushek (1989)].

Há alguma evidência, porém, de que esse resultado pode se dever, em parte, ao tratamento inadequado dispensado pelas estimações por MQO à natureza hierárquica dos dados educacionais. Algumas das variáveis que descrevem a escola são variáveis agregadas das diferentes turmas ou classes que a compõem. Como exemplo, podemos mencionar a escolaridade média dos professores da escola, a proporção de salas de aula arejadas, e assim por diante.

9. Quando os dados permitem, é freqüente a utilização de modelos hierárquicos de três níveis (alunos, turmas, escolas) para estimações desse tipo. Como o número de alunos por turma na base do Saeb é demasiado pequeno para permitir essa desagregação, optou-se por trabalhar somente com dois níveis.

Para ilustrar as possíveis conseqüências de conduzir a análise em um único nível, sem considerar a estrutura dos dados, Bidwell e Kasarda (1980) propõem um modelo teórico no qual o nível socioeconômico dos alunos (X_1) e os recursos escolares (X_2) dessem contribuições da mesma magnitude na explicação do desempenho obtido pelos alunos (Y).¹⁰ A partir desse modelo teórico, os autores simulam uma “pseudobase de dados” de 2.500 alunos e atribuem a cada aluno valores individuais para o nível socioeconômico, assim como valores para os recursos escolares. Os principais resultados das simulações do estudo — que se encontram reproduzidos no Gráfico 1 — indicam um viés sistemático de subestimação do efeito da variável agregada. No caso dos estudos citados anteriormente, esse resultado corresponderia à subestimação das variáveis ligadas às escolas e aos professores.

A introdução de modelos hierárquicos lineares, em substituição à estimação por MQO, poderia alterar tais resultados. Bryk e Raudenbush (1986) explicam que,

“pela sua própria natureza, perguntas a respeito do efeito das escolas sobre o desempenho dos alunos recaem em uma investigação de relações hierárquicas. Tal investigação envolve a procura de associações estatísticas entre os fatores escolares e as variáveis no nível do aluno. Os recentes avanços na teoria estatística dos modelos hierárquicos lineares fornecem ferramentas apropriadas para modelar efeitos entre as escolas e dentro delas.”

GRÁFICO 1
RESULTADOS DAS ANÁLISES

Agrupamento aleatório		Parcialmente por X_1	
i.	X_1 → 0,221 → Y	i.	X_1 → 0,221 → Y
	X_2 → 0,213 → Y		X_2 → 0,213 → Y
ii.	X_1 → 0,262 → Y	ii.	X_1 → 0,262 → Y
	\bar{X}_2 → 0,038 → Y		\bar{X}_2 → 0,014 → Y
iii.	\bar{X}_1 → 0,031 → Y	iii.	\bar{X}_1 → 0,103 → Y
	X_2 → 0,256 → Y		X_2 → 0,248 → Y
iv.	\bar{X}_1 → 0,032 → Y	iv.	\bar{X}_1 → 0,116 → Y
	\bar{X}_2 → 0,039 → Y		\bar{X}_2 → 0,018 → Y

Nota: Coeficientes estimados para o efeito de cada variável elaborada (X_1 e X_2) sobre Y , quando a variável indicada por uma barra superior é agregada. A pseudobase de dados foi elaborada de forma a gerar os valores “verdadeiros” em (i). A primeira coluna refere-se a simulações em que a agregação da variável com barra é feita de forma aleatória. A segunda coluna refere-se a simulações nas quais a agregação é estratificada de acordo com os valores de X_1 .

10. As variâncias de X_1 , X_2 e Y assumiram valores tipicamente reportados em pesquisas educacionais.

O propósito desse tipo de análise é considerar simultaneamente o impacto das características dos alunos e das escolas no desempenho do aluno, levando em conta também que alunos que pertencem à mesma escola não são independentes uns dos outros.

No restante deste artigo, elaboramos um modelo hierárquico linear “de baixo para cima”, ou seja, partindo do modelo mais simples possível até chegar ao modelo completo, em que são identificadas as características escolares capazes de tornar a escola brasileira mais eficaz e equitativa. Na Seção 3, apresentamos brevemente o modelo incondicional e um modelo com todas as características individuais disponíveis para cada aluno. Nessa seção, buscamos respostas para três perguntas:

- a) Que proporção da variância dos rendimentos nos testes do Saeb se dá entre escolas, e que parte se deve à composição da clientela, *versus* as características da própria escola?
- b) Qual o efeito do nível socioeconômico do aluno sobre o seu desempenho?
- c) O grau de associação entre o nível socioeconômico do aluno e o seu desempenho é o mesmo em todas as escolas?

Na Seção 4, expandimos o modelo para incluir variáveis escolares e analisamos os resultados referentes a essas variáveis.

3 UM MODELO HIERÁRQUICO DO RENDIMENTO ESCOLAR

O modelo hierárquico mais simples é conhecido como modelo incondicional, que representa o desempenho do aluno como a soma do desempenho médio da escola (β_{0j}) e de um termo de erro (r_{ij}), que expressa a diferença entre cada observação e a previsão do modelo. Pressupõe-se que esse resíduo siga uma distribuição normal com média zero e variância constante σ^2 .

Como antes explicado, os modelos hierárquicos lineares realizam uma regressão em cada escola. Naturalmente, o grau de confiabilidade das estimativas cresce à medida que aumenta o número de alunos observados por escola. O Saeb tem como característica testar cada aluno em apenas uma disciplina. Assim, para tornar nossas estimativas mais confiáveis, consideramos todos os alunos presentes na amostra. Nesse intuito, foram incluídas quatro variáveis indicadoras, uma para cada disciplina, conforme a equação (2). Note-se que a inclusão dessas variáveis ajusta as diferenças de escala entre as disciplinas.

$$Y_{ij} = \beta_{0j} + \beta_{1j} * dcie_{ij} + \beta_{2j} * dgeo_{ij} + \beta_{3j} * dhit_{ij} + \beta_{4j} * dport_{ij} + r_{ij} \quad (2)$$

No nível da escola, representa-se o desempenho médio da escola em matemática (β_{0j}) como função da média geral de todas as escolas (γ_{00}), mais um componente aleatório (u_{0j}). Em um primeiro momento, considera-se que o desempenho médio das escolas nas demais disciplinas não varia de escola para escola, e, portanto, estas não apresentam um termo de erro aleatório.¹¹ A equação de nível 2 pode ser assim expressa:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40}\end{aligned}\tag{3}$$

Novamente, supõe-se que u_{0j} se comporta como um erro aleatório, distribuído normalmente com média zero e variância τ_{00} . Os valores estimados desses coeficientes encontram-se na Tabela 3.

A Tabela 3 contém dois resultados de interesse. Primeiro, a variância de β_{0j} ao redor de γ_{00} é alta ($\hat{\tau}_{00} = 733,9$) e estatisticamente significativa. Desse modo, a

TABELA 3
MODELO INCONDICIONAL^a

Efeitos fixos	Coefficiente	Erro-padrão	Teste-t	Valor-p
Média das escolas em matemática (γ_{00})	253,1	0,65	385,9	0,00
Desvio médio de ciências em relação a matemática (γ_{10})	-2,5	0,44	-5,7	0,00
Desvio médio de geografia em relação a matemática (γ_{20})	3,5	0,44	8,1	0,00
Desvio médio de história em relação a matemática (γ_{30})	3,3	0,44	7,6	0,00
Desvio médio de português em relação a matemática (γ_{40})	-14,2	0,44	-32,1	0,00
Efeitos aleatórios	Variância	Graus de liberdade	χ^2	Valor-p
Média da escola (u_{0j})	733,9	2.578	44.727,6	0,00
Efeito do aluno (r_{ij})	1.846,7			

^a A terminologia habitual na literatura de HLM refere-se aos coeficientes estimados como "efeitos fixos" e às estimativas de variância como "efeitos aleatórios". Os termos são utilizados aqui com esta interpretação, e não com aquela associada à econometria de dados em painel.

11. O desempenho médio das escolas nas demais disciplinas é considerado fixo, porque não dispomos de um grande número de alunos em cada escola.

hipótese nula de nenhuma diferença entre os coeficientes de intercepto das escolas é descartada, ou seja, a eficácia das escolas varia significativamente em torno da média nacional. Em segundo lugar, a correlação intraclasse é dada por $\hat{\rho} = \hat{\tau}_{00} / (\hat{\tau}_{00} + \hat{\sigma}^2) = 733,9 / (733,9 + 1.846,7) = 0,28$,¹² ou seja, 28% da variância total do desempenho em matemática ocorrem entre as escolas. Para testar se uma parte expressiva da variância de desempenho dos alunos nas demais disciplinas também ocorre entre escolas, calculamos a correlação intraclasse para cada uma das disciplinas. Em todos os casos, essa correlação excedeu 0,15.¹³ Assim, prosseguimos na análise adotando os modelos hierárquicos lineares na estimação da função de produção educacional brasileira.¹⁴

Os próximos passos a serem tomados buscam entender quais características dos alunos e das escolas tornam as últimas mais eficazes e equitativas. Inicialmente, são incluídas sete variáveis de controle no nível dos alunos: quatro variáveis indicadoras de repetência, cada uma representando o número total de anos que o aluno repetiu; duas variáveis indicadoras de raça; e uma variável relacionada ao sexo dos alunos. A inclusão da variável de repetência é motivada pelo fato de a variável dependente — o desempenho do aluno (Y_{ij}) — ser uma medida única em todo o seu percurso escolar. Assim, a nota obtida no teste-padrão reflete todo o aprendizado durante os anos de sua permanência na escola até o momento do teste.¹⁵ Quando as variáveis de repetência são incluídas, elas espelham — em parte — o percurso escolar dos indivíduos que fizeram o teste. A inclusão das variáveis de sexo e de raça no nível dos alunos tem por objetivo controlar possíveis efeitos do gênero e da etnia sobre o seu desempenho.

O novo modelo de nível 1 incorpora também o efeito do nível socioeconômico no desempenho dos alunos.¹⁶ Esse modelo é representado pela equação (4).

$$\begin{aligned}
 Y_{ij} = & \beta_{0j} + \beta_{1j} * (masc)_{ij} + \beta_{2j} * (dcie)_{ij} + \beta_{3j} * (dgeo)_{ij} + \beta_{4j} * (dhit)_{ij} + \\
 & \beta_{5j} * (dport)_{ij} + \beta_{6j} * (nse)_{ij} + \beta_{7j} * (repete_1)_{ij} + \beta_{8j} * (repete_2)_{ij} + \\
 & \beta_{9j} * (repete_3)_{ij} + \beta_{10j} * (repete_4)_{ij} + \beta_{11j} * (pardo)_{ij} + \\
 & \beta_{12j} * (preto)_{ij} + r_{ij}
 \end{aligned} \quad (4)$$

12. A correlação intraclasse representa a proporção da variância total que corresponde à variância entre as escolas. Os valores tipicamente encontrados para a correlação intraclasse na literatura dos modelos hierárquicos estão entre 0,20 e 0,25.

13. Para obter maiores detalhes, ver Albernaz (2002).

14. A partir do modelo incondicional calculou-se também o índice de fidedignidade das médias escolares, que ficou em 0,808. Esse resultado mostra que a informação utilizada pelo modelo multinível para calcular a estimativa da média escolar provém, em média, 80% da própria escola e 20% de outras escolas.

15. Como comentado na Seção 2, o mais correto seria utilizar uma estimativa de valor agregado para a função de produção educacional. No entanto, ainda não se encontra disponível para pesquisa uma base de dados brasileira que conte com duas medidas de desempenho de um mesmo aluno em dois pontos no tempo.

16. A descrição dessa variável encontra-se na Tabela 2.

O modelo de nível 2, por sua vez, incorpora o efeito do nível socioeconômico médio da clientela da escola ($nsemédio$)_{.j} sobre o desempenho médio da escola (β_{0j}) e sobre o efeito do nível socioeconômico dos alunos (β_{6j}). A equação de nível 2 pode ser assim expressa:

$$\begin{aligned}
 \beta_{0j} &= \gamma_{00} + \gamma_{01} * (nsemédio)_{.j} + u_{0j} \\
 \beta_{1j} &= \gamma_{10} \\
 \beta_{2j} &= \gamma_{20} \\
 \beta_{3j} &= \gamma_{30} \\
 \beta_{4j} &= \gamma_{40} \\
 \beta_{5j} &= \gamma_{50} \\
 \beta_{6j} &= \gamma_{60} + \gamma_{61} * (nsemédio)_{.j} + u_{6j} \\
 \beta_{7j} &= \gamma_{70} \\
 \beta_{8j} &= \gamma_{80} \\
 \beta_{9j} &= \gamma_{90} \\
 \beta_{10j} &= \gamma_{100} \\
 \beta_{11j} &= \gamma_{110} \\
 \beta_{12j} &= \gamma_{120}
 \end{aligned} \tag{5}$$

Os efeitos fixos e aleatórios para o modelo de verificação da eficácia e equidade entre as escolas encontram-se na Tabela 4. Os dois resultados centrais desta tabela referem-se ao efeito do nível socioeconômico sobre o desempenho do aluno. O valor médio dessa variável em cada escola está forte e significativamente relacionado com o desempenho médio do aluno ($\gamma_{01} = 25,5$, $t = 64,20$). Mesmo controlando por esse efeito de grupo, o nível socioeconômico individual influencia também o desempenho de cada aluno dentro da escola. No nível do aluno, o nível socioeconômico contribui de forma estatisticamente significativa, mesmo que em menor grau, para o seu desempenho ($\gamma_{60} = 1,2$, $t = 4,35$).¹⁷

Controlando pelo nível socioeconômico dos alunos, a Tabela 4 revela ainda a existência de três outros fatores individuais com efeitos estatisticamente signifi-

17. Conforme apresentado na Tabela 2, o nível socioeconômico do aluno varia de -2,84 até 9,20, enquanto o nível socioeconômico médio da clientela varia de -2,11 até 2,98. Observa-se, ainda, que, quando se inclui a etnia do aluno no modelo, há uma pequena queda no efeito do nível socioeconômico do aluno sobre o seu desempenho.

TABELA 4
MODELO DE VERIFICAÇÃO DA EFICÁCIA E EQUIDADE ENTRE AS ESCOLAS

Efeitos fixos	Coefficiente	Erro-padrão	Teste-t	Valor-p
Média da escola (β_{0j})				
Intercepto (γ_{00})	260,7	0,47	552,91	0,00
Nsemédio (γ_{01})	25,5	0,40	64,20	0,00
Sexo masculino (γ_{10})	3,5	0,28	12,35	0,00
Média das escolas em ciências (γ_{20})	-2,4	0,43	-5,50	0,00
Média das escolas em geografia (γ_{30})	3,5	0,43	8,10	0,00
Média das escolas em história (γ_{40})	3,6	0,43	8,20	0,00
Média das escolas em português (γ_{50})	-13,9	0,43	-32,17	0,00
Inclinação desempenho-nse (β_{6j})				
Intercepto (γ_{60})	1,2	0,27	4,35	0,00
Nsemédio (γ_{61})	-1,9	0,33	-5,74	0,00
Aluno com 1 ano de repetência (γ_{70})	-16,0	0,35	-45,80	0,00
Aluno com 2 anos de repetência (γ_{80})	-20,0	0,45	-43,34	0,00
Aluno com 3 anos de repetência (γ_{90})	-21,0	0,67	-31,25	0,00
Aluno com 4 anos de repetência (γ_{100})	-22,8	1,01	-22,53	0,00
Aluno pardo/mulato (γ_{110})	-2,5	0,32	-8,01	0,00
Aluno preto (γ_{120})	-7,5	0,58	-12,79	0,00
Efeitos aleatórios	Variância	Graus de liberdade	χ^2	Valor-p
Média da escola (u_{0j})	143,5	2.575	12.614,0	0,00
Inclinação do nse (u_{6j})	34,3	2.575	2.839,2	0,00
Efeito do aluno (r_{ij})	1.744,2			

cativos sobre o desempenho estudantil: histórico escolar, gênero e raça. No que diz respeito ao histórico escolar, enquanto os alunos com um ano de repetência apresentam, em média, desempenho inferior em 16 pontos ($\gamma_{70} = -16,0$), os alunos com dois anos de repetência apresentam um desempenho médio ainda inferior ($\gamma_{80} = -20,0$) e os que repetiram três e quatro vezes ($\gamma_{90} = -21,0$, $\gamma_{100} = -22,8$). Esse forte impacto negativo da repetência sobre o desempenho dos alunos confirma a importância de tentarmos controlar pelo estoque de capital humano com o

qual o aluno chega ao teste, de forma a não atribuírmos erroneamente ao ensino recente problemas advindos de trajetórias passadas.¹⁸

Com relação à raça, a tabela mostra que os alunos pardos ou mulatos têm desempenho inferior ao observado para os alunos brancos¹⁹ em 2,5 pontos. Essa diferença é ainda maior no caso dos alunos pretos ($\gamma_{120} = -7,5$). Esse efeito negativo da cor sobre o rendimento escolar, mesmo após o controle pelo nível socioeconômico, constitui um resultado altamente preocupante para os interessados em reduzir a desigualdade de oportunidades no Brasil. O negro brasileiro parece não só ter menos chance de estar na escola como também, quando nela ingressa e aí logra permanecer, ter um desempenho pior do que seus colegas brancos, mesmo controlando pelo nível socioeconômico.

O tema da equidade também é abordado quando avaliamos a maneira pela qual o nível socioeconômico do aluno afeta o seu desempenho. Repare que este conceito está intimamente ligado à eficácia da escola (β_{0j}), pois só faz sentido falar sobre equidade quando existe alguma variável que contribui, em média, para um melhor desempenho da escola. Ainda de acordo com a Tabela 4, o incremento de uma unidade no nível socioeconômico médio da clientela da escola reduz o impacto do nível socioeconômico do aluno no seu desempenho em 1,9 ponto.

A Tabela 4 também apresenta informações sobre os efeitos aleatórios. A variância residual entre as escolas ($\hat{\tau}_{00} = 143,5$) caiu para 19,6% em relação ao modelo incondicional, onde $\hat{\tau}_{00} = 733,9$, o que implica que 80,4% da variância observada entre as escolas no desempenho médio de seus alunos derivam da composição social interna e da seletividade da escola. Ainda assim, a variância que restou entre as escolas é estatisticamente significativa. Ou seja, há diferenças entre as escolas — tanto no que tange à eficácia, quanto no que diz respeito à equidade — que não são explicadas por diferenças relativas a características da clientela, tais como composição socioeconômica, racial ou de gênero. A próxima seção utiliza dados sobre as escolas propriamente ditas para investigar se parte dessa variância não-explicada deve-se a características da escola em si ou de seus professores.

4 O PAPEL DAS CARACTERÍSTICAS E PRÁTICAS ESCOLARES

Como vimos na seção anterior, uma proporção significativa da variância total dos resultados no teste do Saeb de 1999 era entre médias escolares, o que amplamente respaldava a visão de que as escolas brasileiras não são estatisticamente

18. Naturalmente, é plausível que as variáveis de repetência capturem também algum efeito de características não-observáveis dos alunos, com elas correlacionadas.

19. Nessa categoria também se enquadram os alunos amarelos, que têm desempenho médio semelhante ao dos alunos brancos, bem como os alunos indígenas. Esses últimos representam menos de 3% da amostra.

idênticas em termos de preparo para o teste. Vimos também que, dessa porção da variância existente entre escolas, cerca de 80% são explicados por diferenças no nível socioeconômico médio da clientela. Ocorre, claramente, um fenômeno de estratificação (*sorting*) dos alunos por critérios econômicos, que reforça o impacto positivo que essa variável tem em nível individual.²⁰

Não obstante a importância quantitativa desse efeito, observamos ainda que a variância restante, mesmo após o controle pelo nível socioeconômico, permanecia estatística e economicamente significativa. Nesta seção, incluímos no modelo anterior as variáveis escolares disponíveis na base de dados, para testar o poder explicativo destas no desempenho dos estudantes. As variáveis escolares estão classificadas em três grupos. O primeiro refere-se a itens de infra-estrutura da escola, o segundo analisa o impacto da disponibilidade dos recursos financeiros na escola, enquanto o terceiro está relacionado às práticas escolares.

No intuito de verificar o impacto de cada variável escolar na função de produção educacional, foi introduzida na modelagem do intercepto (β_{0j}) e da inclinação (β_{6j}) uma variável de cada vez. Embora muitas variáveis se mostrassem importantes na modelagem do intercepto, somente duas foram (alternativamente) relevantes para a modelagem da inclinação: o nível de escolaridade e o salário dos professores. O próximo modelo reporta como essas variáveis influenciam o desempenho do aluno.

Observa-se que o modelo de nível 1 continua a ser representado pela equação (4), enquanto o modelo de nível 2 passa a ser representado pela equação (6):²¹

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + \gamma_{01} * (particular)_{.j} + \gamma_{02} * (saa)_{.j} + \gamma_{03} * (rp)_{.j} + \\ & \gamma_{04} * (irf)_{.j} + \gamma_{05} * (faltaalu)_{.j} + \gamma_{06} * (nesxpr)_{.j} + \\ & \gamma_{07} * (nsemédio)_{.j} + \gamma_{08} * (prop_rep)_{.j} + u_{0j} \end{aligned}$$

20. Um modelo econômico de estratificação por critério duplo — de mérito acadêmico e nível socioeconômico — é encontrado em Fernandez (1997).

21. Foram testadas as interações das variáveis e a grande maioria revelou-se não-significativa. Desse modo, optou-se por não reportá-las.

$$\begin{aligned}
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30} \\
\beta_{4j} &= \gamma_{40} \\
\beta_{5j} &= \gamma_{50} \\
\beta_{6j} &= \gamma_{60} + \gamma_{61} * (nexpr)_{.j} + \gamma_{62} * (nsemédio)_{.j} + \gamma_{63} * (prop_rep)_{.j} + u_{6j} \quad (6) \\
\beta_{7j} &= \gamma_{70} \\
\beta_{8j} &= \gamma_{80} \\
\beta_{9j} &= \gamma_{90} \\
\beta_{10j} &= \gamma_{100} \\
\beta_{11j} &= \gamma_{110} \\
\beta_{12j} &= \gamma_{120}
\end{aligned}$$

Os resultados da estimação do modelo (4) - (6) encontram-se na Tabela 5.²² Na literatura das funções de produção da educação, a variável escolar que se mostrou significativa um maior número de vezes para explicar a variância do desempenho dos alunos foi a experiência do professor. No contexto educacional brasileiro, essa variável não parece ser relevante, pois em nenhum momento foi significativa: nem na modelagem do intercepto, nem na da inclinação. Em contrapartida, a variável “particular”, que indica se a escola pertence à rede pública ou particular, mostrou-se altamente significativa na modelagem da eficácia. Controlando-se por todas as variáveis escolares e individuais, bem como pelo nível socioeconômico médio dos alunos, a escola particular agrega ainda 6,1 pontos à média dos resultados de seus alunos.

Quando a variável “particular” foi introduzida na modelagem do parâmetro de equidade (β_6), apresentou uma estimativa pontual próxima de zero, além de ter se mostrado não-significativa. Desse modo, essa variável foi omitida da especificação apresentada.

22. A correlação entre o nível socioeconômico médio e algumas variáveis escolares pode redundar em subestimação dos coeficientes estimados para as variáveis escolares. Em sentido estrito, o efeito efetivo das variáveis escolares pode ser equacionado da seguinte forma: é maior do que o representado pelos coeficientes de regressão estimados em modelos que implementam controle de nível socioeconômico médio no nível 2; e é menor que o representado pelos coeficientes de regressão estimados em modelos que não implementam controle de nível socioeconômico médio no nível 2. No entanto, deve-se considerar que a já discutida inexistência de medida de valor agregado e a conseqüente substituição do controle por proficiência prévia por um controle menos estrito (repetência) redundam em tendência de superestimação do efeito das variáveis escolares. A existência dessa tendência nos estimula a tomar a decisão mais prudente, que é a de apenas reportar efeitos de variáveis escolares estimados através de modelos que implementam controle de nível socioeconômico médio.

TABELA 5
MODELO EDUCACIONAL COM NÍVEL DE ESCOLARIDADE DO PROFESSOR

Efeitos fixos	Coefficiente	Erro-padrão	Teste-t	Valor-p
Média da escola (β_{0j})				
Intercepto (γ_{00})	266,2	1,58	168,68	0,00
Particular (γ_{01})	6,1	1,09	6,00	0,00
Sala de aula arejada (γ_{02})	3,5	0,91	3,91	0,00
Ruído prejudica (γ_{03})	-1,2	0,69	-1,79	0,07
Insuficiência de recursos financeiros (γ_{04})	-5,1	1,05	-4,86	0,00
Falta aluno (γ_{05})	-4,3	1,19	-3,60	0,00
Nível de escolaridade dos professores (γ_{06})	2,9	1,17	2,50	0,01
Nsemédio (γ_{07})	18,8	0,69	26,34	0,00
Proporção de repetentes (γ_{08})	-12,5	1,95	-6,43	0,00
Sexo masculino (γ_{10})				
	3,5	0,29	12,40	0,00
Média das escolas em ciências (γ_{20})				
	-2,6	0,44	-5,84	0,00
Média das escolas em geografia (γ_{30})				
	3,4	0,44	7,78	0,00
Média das escolas em história (γ_{40})				
	3,5	0,44	7,97	0,00
Média das escolas em português (γ_{50})				
	-14,0	0,44	-31,95	0,00
Inclinação desempenho - nse (β_{6j})				
Intercepto (γ_{60})	2,8	0,75	3,75	0,00
Nível de escolaridade dos professores (γ_{61})	2,1	1,06	2,03	0,04
Nsemédio (γ_{62})	-2,8	0,47	-6,09	0,00
Proporção de repetentes (γ_{63})	-4,0	1,56	-2,54	0,01
Aluno com 1 ano de repetência (γ_{70})				
	-15,5	0,36	-43,48	0,00
Aluno com 2 anos de repetência (γ_{80})				
	-19,3	0,47	-41,33	0,00
Aluno com 3 anos de repetência (γ_{90})				
	-20,3	0,68	-29,66	0,00
Aluno com 4 anos de repetência (γ_{100})				
	-21,9	1,04	-21,06	0,00
Aluno pardo/mulato (γ_{110})				
	-2,4	0,32	-7,52	0,00
Aluno preto (γ_{120})				
	-7,4	0,59	-12,42	0,00
Efeitos aleatórios				
	Variância	Graus de liberdade	χ^2	Valor-p
Média da escola (u_{0j})	126,9	2.475	11.238,8	0,00
Inclinação do nse (u_{6j})	31,9	2.480	2.715,2	0,00
Efeito do aluno (r_{ip})				
	1.745,5			

A partir da Tabela 5, observa-se que o desempenho médio das escolas depende de uma série de fatores escolares, depois de controlado o efeito do nível socioeconômico médio de seus alunos e da rede de ensino a que ela pertence e da proporção de repetentes. Em comparação com modelos intermediários (não reportados), a introdução desses novos fatores reduziu os efeitos antes atribuídos à rede de ensino e ao nível socioeconômico médio dos alunos, cujos coeficientes passam a ser, respectivamente, $\gamma_{01} = 6,1$, $t = 6,00$ e $\gamma_{07} = 18,8$, $t = 26,34$.

Um resultado que chama a atenção é o impacto grande e significativo da proporção de repetentes sobre o desempenho médio das escolas ($\gamma_{08} = -12,5$, $t = -2,54$). Esse impacto é significativo mesmo após o controle do nível socioeconômico médio dos alunos da escola e da rede de ensino, o que significa que a experiência prévia de repetência está associada a piores resultados educacionais, independentemente do tipo de escola que o aluno frequenta.²³

O primeiro grupo de variáveis escolares, referente às condições de infraestrutura física, está representado pelas variáveis sala de aula arejada e nível de ruído das salas de aula. Quanto melhor for a infra-estrutura escolar, melhor será o ambiente de estudo para os alunos, o que irá favorecer o seu desempenho. Isso pode ser confirmado na Tabela 5, pois, quando a escola possui salas de aula arejadas, o desempenho médio dos alunos da escola aumenta, $\gamma_{02} = 3,5$ pontos, e quando o nível de ruído da escola prejudica os alunos e professores, o desempenho cai, em média, 1,2 ponto.

A variável que está relacionada ao segundo grupo mede a insuficiência de recursos financeiros. Constatou-se que quando a escola apresenta problemas de caráter financeiro, o desempenho do aluno cai, em média, 5,1 pontos. O terceiro e último grupo está relacionado às práticas escolares e trata de questões como a incidência de falta às aulas, por parte dos alunos, e o nível de escolaridade do professor.²⁴ A Tabela 5 descreve o importante impacto do nível de escolaridade do professor sobre o desempenho médio dos alunos ($\gamma_{06} = 2,9$).²⁵ Como é de se esperar, quanto maior a incidência de falta às aulas, por parte dos alunos, menor é o desempenho médio deles ($\gamma_{05} = -4,3$). Na realidade, essa variável também

23. É natural imaginar que esse resultado possa também refletir a existência de alguma correlação serial temporal da variável dependente: as escolas piores podem ter mais repetentes porque elas já eram ruins quando esses alunos para lá foram estudar.

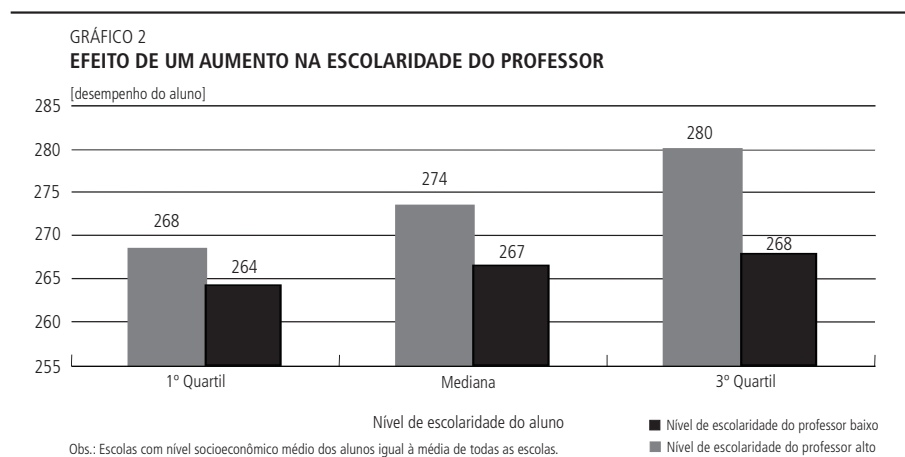
24. Como está descrito na Tabela 2, os valores tomados por essa variável oscilam entre 0 e 3, em que 3 se refere aos professores que possuem formação de ensino superior.

25. Quando se inclui a variável salário do professor na modelagem de β_{0j} e β_{6p} , o nível educacional do professor torna-se insignificante para explicar tanto o melhor desempenho do aluno, quanto a relação entre seu nível socioeconômico e seu desempenho. Substituiu-se, então, a variável nível salarial por nível educacional do professor. A partir desse novo modelo chegou-se à conclusão de que o salário do professor impacta o desempenho dos alunos de forma semelhante ao nível de escolaridade do professor. Enquanto um salário maior do professor se traduz em um maior desempenho médio da escola, à medida que o salário do professor aumenta, o grau de associação entre o nível socioeconômico do aluno e seu desempenho se torna mais forte.

capta, mesmo que parcialmente, o grau de interesse dos alunos pelas aulas. Assim, é razoável observar-se uma diminuição no desempenho médio dos alunos quando estes não se mostram interessados nas aulas.

A Tabela 5 retrata também qual o impacto do nível socioeconômico médio dos alunos, do nível de escolaridade do professor e da proporção de repetentes na escola sobre a maneira como o nível socioeconômico do aluno afeta a sua proficiência (β_{6j}). Observa-se que um maior nível socioeconômico médio dos alunos faz com que o nível socioeconômico familiar do aluno afete menos o seu desempenho²⁶ ($\gamma_{62} = -2,8$). Por sua vez, um maior nível de escolaridade do professor tem o efeito oposto: quanto mais instruído for o professor, maior o impacto do nível socioeconômico do aluno no seu desempenho ($\gamma_{61} = 2,1$).

Desse modo, se, por um lado, um maior nível de escolaridade do professor faz com que a escola seja mais eficaz — aumenta o desempenho médio dos alunos que a frequentam —, por outro, torna a escola menos equitativa — aumenta a importância do nível socioeconômico individual do aluno para o seu desempenho. Uma outra maneira de ver o mesmo efeito encontra-se ilustrada no Gráfico 2, em que o aumento na escolaridade dos professores beneficia todos os alunos da escola.²⁷ Esse benefício é crescente com o nível socioeconômico familiar do aluno, de forma que professores mais qualificados contribuem para o aprendizado de todos, mas de forma mais efetiva para o dos alunos de nível socioeconômico mais elevado.²⁸



26. Quer dizer, o resultado da Tabela 4 permanece válido neste modelo mais geral.

27. O Gráfico 2 apresenta o desempenho de alunos apenas em escolas com nível socioeconômico médio entre $-0,4$ e $0,4$, de forma a controlar para o efeito do nível socioeconômico médio da escola.

28. No Gráfico 2, a distribuição do nível socioeconômico familiar dentro da escola é representada por três pontos: o primeiro quartil, a mediana e o terceiro quartil.

Um último exercício foi realizado, no sentido de testar a robustez dos parâmetros estimados à omissão da variável “rede” do modelo. Assim, reestimamos o modelo (4) - (6), excluindo a variável rede, e reexaminamos todas as outras variáveis disponíveis, a fim de verificar qualquer mudança de significância ou magnitude nos coeficientes. A única variável não significativa a 5% que passou a sê-lo foi a não-disponibilidade de professores para lecionar determinadas matérias (resposta dada pelo diretor da escola). Quando a variável “rede” é omitida, a não-disponibilidade de professores passa a afetar negativamente o desempenho do aluno, sugerindo que uma das razões por trás da menor eficácia da escola pública poderia ser um menor sucesso em garantir a disponibilidade efetiva dos professores necessários para a completa cobertura do currículo.

5 CONCLUSÕES

Através da aplicação de um método estatístico apropriado (a estimação de um modelo hierárquico linear) a uma base de dados relativamente nova (os resultados dos testes padronizados do Saeb de 1999) tentamos, neste artigo, contribuir para uma melhor compreensão dos determinantes da qualidade do ensino fundamental no Brasil. Os resultados encontrados agrupam-se, de forma bastante natural, em duas categorias: a primeira refere-se ao papel das características individuais e familiares dos próprios estudantes, e a segunda ao papel das variáveis escolares e dos professores.

Na primeira categoria, o resultado central é a enorme importância não apenas do nível socioeconômico da família como determinante do desempenho do aluno, mas também do nível socioeconômico médio da clientela da escola que ele frequenta, que, quanto mais alto, melhor é o desempenho médio dos alunos da escola. E mesmo após o controle pelo nível socioeconômico médio da clientela, quanto maior é o nível socioeconômico individual, melhor o seu rendimento escolar na escola.²⁹ Mas esse efeito positivo do nível socioeconômico individual não é igual em todas as escolas. Na verdade, quanto maior o nível socioeconômico médio da clientela das escolas, menor o efeito das condições familiares de cada aluno. Esse resultado da modelagem sintetiza dois fatores relevantes distintos. Primeiro, como escolarização envolve dinâmicas sociais, o efeito de agrupamento mostra-se relevante para a produção dos resultados educacionais, de modo que enquanto estudantes de nível socioeconômico baixo tendem a ser duplamente punidos — devido às condições de sua família e à maior probabilidade de estudar em escolas de clientela com nível socioeconômico médio baixo — estudantes de nível socioeconômico alto têm seus resultados educacionais positivamente influen-

29. Esse resultado também é encontrado em outros países, tanto na América Latina [ver Mizala e Romaguera (1998) e Mizala, Romaguera e Reinaga (1999)], como nos Estados Unidos [ver Hanushek (1989)].

ciados pelo efeito de grupo. O segundo fator a ser considerado é que, em parte, a variável nível socioeconômico médio pode capturar efeitos de variáveis escolares omitidas, se estas forem com ela correlacionadas.

O pior desempenho do aluno negro (preto ou pardo), mesmo após o controle pelo nível socioeconômico é preocupante. Supondo que erros de medida na variável nível socioeconômico sejam ortogonais à variável raça, a significância estatística desse resultado depõe contra a existência de igualdade de oportunidades na escola brasileira. Trabalhos anteriores como, por exemplo, o de Henriques (2001) já apontavam que os negros têm menores taxas de acesso e freqüência à escola. O presente resultado sugere que, mesmo dentro da escola e controlando pelo seu nível socioeconômico, o aluno negro tem um pior desempenho de aprendizado. Em um contexto onde políticas afirmativas começam a ser debatidas no país, cremos ser necessária uma investigação cuidadosa e detalhada dos aspectos raciais da prática educacional nas escolas brasileiras. Existe alguma forma de discriminação, ainda que sutil e talvez inconsciente, na alocação do aluno negro à sala de aula da parte do professor ou em sua atenção individual? Existe algum problema de relevância cultural do currículo ou do material didático que possa torná-los menos interessantes ou inapropriados para alunos de minorias raciais? Essas são questões que vão muito além do escopo deste trabalho, mas que os presentes resultados sugerem como temas relevantes para pesquisa futura.

Passando à categoria dos resultados sobre as variáveis escolares, cabe ressaltar que a significância estatística dos efeitos de um bom número delas sobre o desempenho do aluno contrasta com os resultados usuais na literatura econômica americana e latino-americana. Isso pode se dever a uma diferença legítima entre o Brasil e os outros países estudados, ou a diferenças metodológicas, principalmente devido a uma redução no viés de agregação que pode ter levado à subestimação dos coeficientes dessas variáveis em análises anteriores. Seja como for, tanto a qualidade do professor — medida principalmente por seu nível de escolaridade — como a qualidade da estrutura física da escola — recursos financeiros, salas arejadas e silenciosas — contribuem para um melhor desempenho de seus alunos.³⁰

Cabe notar que esse resultado não só confirma que escolas melhores levam a melhores resultados — o que não chegaria a surpreender — como também mostra que essa variância de qualidades escolares pode ser efetivamente observada na amostra das escolas brasileiras em operação em 1999. Isso quer dizer que há

30. Mesmo controlando por todos esses fatores explicitamente, assim como para a composição étnica e socioeconômica da clientela, as escolas públicas apresentaram um desempenho inferior estatisticamente significativo em relação às escolas particulares no Brasil. Caberia investigar se há aspectos de gestão, condições estruturais de ensino e/ou incentivos internos aos funcionários da rede pública que podem ser responsáveis por esse pior desempenho. Esse resultado também se coaduna com modelos da economia política do financiamento à educação no Brasil [ver Ferreira (2001)].

escolas no Brasil onde as crianças estão aprendendo menos do que deveriam, por causa da insuficiência de recursos financeiros, da insuficiência de professores, de sua baixa escolaridade e de salas barulhentas e/ou abafadas. Não é preciso um grande esforço intelectual para discernir as implicações desse fato para a política pública, com conseqüências tanto para a eficácia quanto para a igualdade de oportunidades em nosso sistema educacional.

Outro resultado com possíveis implicações para a política educacional é o fato de incrementos na escolaridade do professor contribuírem para um melhor desempenho de todos os alunos, mas de forma crescente de acordo com o nível socioeconômico destes. Claramente, o aumento da escolaridade do professor, por aumentar a eficácia das escolas, já é, por si, desejável. Entretanto, para reduzir a desigualdade de oportunidades educacionais entre famílias mais e menos afortunadas, as autoridades públicas deveriam tentar desenvolver políticas educacionais e pedagógicas voltadas para identificar e retificar as razões específicas por trás do pior desempenho dos alunos mais pobres. Tais políticas poderiam incluir uma disponibilidade maior do professor fora do horário de aula, no atendimento aos alunos; programas de incentivo ao uso da biblioteca; e disponibilidade de acompanhamento psicológico, entre outras sugestões.³¹ A investigação do mérito de cada sugestão foge ao escopo deste estudo, no qual apenas constatamos a necessidade de tais políticas para evitar que aumentos na eficácia escolar sejam acompanhados por aumentos na desigualdade de oportunidades na escola.

ABSTRACT

Using Ministry of Education standardized test score data (from Saeb) and hierarchical linear models—so as to reduce aggregation bias in the estimation of the effects of school variables—this paper estimates an educational production function for primary schooling performance in Brazil in 1999. As elsewhere, we find that family and individual characteristics are fundamental determinants of educational achievement. In particular, 80% of the interschool variation in test scores can be attributed to differences in the average socioeconomic group of its students. Additionally, both socioeconomic level and race affect individual performance within schools. However, and in contrast with much of the international literature, we also find significant effects of school variables on student performance. Better educated teachers, better physical facilities and private ownership of schools are all jointly and severally associated with higher test scores.

31. É óbvio que cada uma dessas políticas teria custos. O que se aponta aqui é que elas também teriam benefícios, com os quais os custos devem ser comparados.

BIBLIOGRAFIA

- ALBERNAZ, A. *Determinantes do desempenho educacional no ensino fundamental brasileiro*. PUC-Rio, 2002 (Tese de Mestrado em Economia).
- ARROW, K. J. The economic implications of learning by doing. *Review of Economic Studies*, v. 29, p. 155-173, 1962.
- BARBOSA, M., FERNANDES, C. A escola brasileira faz diferença?: uma investigação dos efeitos da escola na proficiência em matemática dos alunos da quarta série. In: FRANCO, C. (ed.). *Promoção, ciclos e avaliação na educação*. Porto Alegre: ArtMed, 2001.
- BARROS, R. P. de, HENRIQUES, R., MENDONÇA, R. Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil. *Desigualdade e Pobreza no Brasil*, IPEA, Cap. 14, p. 405-423, 2000.
- BARROS, R. P. de, MENDONÇA, R., SANTOS, D. *Determinantes do desempenho educacional no Brasil*. IPEA, 1999, mimeo.
- BECKER, G., MURPHY, K., TAMURA, R. Human capital, fertility, and economic growth. *Journal of Political Economy*, v. 98, p. s12-s37, 1990.
- BIDWELL, E., KASARDA, D. Conceptualizing and measuring the effects of school and schooling. *Journal of Education*, p. 401-430, 1980.
- BRYK, S., RAUDENBUSH, W. A hierarquical model for studying school effects. *Sociology of Education*, v. 59, p. 1-17, 1986.
- . *Hierarquical linear models: applications and data analysis method*. Newbury Park: Sage, 1992.
- BUSSAB, W., SILVA, P., ANDRADE, D. *Saeb/99, relatório sobre a amostra, informe metodológico*. Inep, Ministério da Educação, 2000.
- COLEMAN, S. *Equality of educational opportunity*. U.S., Washington D.C.: Office of Education, 1966.
- FERNANDEZ, R. *To each according to?: markets, tournaments, and the matching problem with borrowing constraints*. New York University, 1997, mimeo.
- FERREIRA, F. H. G. Education for the masses?: the interaction between wealth, educational and political inequalities. *Economics of Transition*, v. 9, n. 2, p. 533-552, 2001.
- FLETCHER, R. *À procura do ensino eficaz*. PNUD/MEC/Saeb, 1997 (Relatório de Pesquisa).
- FRANCO, C., MANDARINO, M., ORTIGÃO, M. I. Projeto pedagógico de escola promove qualidade e equidade em educação? *Revista Undime-RJ*, v. VII, n. 2, p. 30-46, 2001.
- GOLDSTEIN, H. *Multilevel statistical models*. Kendall's Library of Statistics, 1995.
- HANUSHEK, E. The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, v. 24, p. 1.141-1.177, 1986.
- . The impact of differential expenditures on school performance. *Educational Researcher*, p. 45-52, 1989.
- HENRIQUES, R. *Desigualdade racial no Brasil: evolução das condições de vida na década de 90*. Rio de Janeiro: IPEA, 2001 (Texto para Discussão, 807).

- LANGONI, C. *Distribuição de renda e desenvolvimento econômico do Brasil*. Rio de Janeiro: Expressão e Cultura, 1973.
- LUCAS, R. On the mechanics of economic development. *Journal of Monetary Economics*, v. 22, p. 3-42, 1988.
- MATTAR, N. Análise crítica dos estudos de estratificação socioeconômica de ABA-Abipeme. *Revista de Administração*, São Paulo, v. 30, n. 1, p. 57-74, 1995.
- MINCER, J. Investment in human capital and personal income distribution. *Journal of Political Economy*, v. 66, p. 281-302, 1958.
- MINISTÉRIO DA EDUCAÇÃO. *Relatório do Saeb 1999*. Diretoria de Avaliação da Educação Básica, Inep, 1999 (<http://www.inep.gov.br>).
- MIZALA, A., ROMAGUERA, P. *Desempeño escolar y eleccion de colegios: la experiencia chilena*. Universidad de Chile, 1998 (Documento de Trabajo, 36, Serie Economía).
- MIZALA, A., ROMAGUERA, P., REINAGA, T. *Determinants of student achievement and school performance in Bolivia*. CD-ROM Lacea 2000, 1999.
- MOLENAAR, I. W. Nonparametric models for polytomous responses. In: LINDEN, W. J. van der, HAMBLETON, R. K. (orgs.). *Handbook of Modern Item Response Theory*. New York: Springer, p. 369-380, 1997.
- SUMMERS, A., WOLFE, B. Do schools make a difference? *The American Economic Review*, v. 67, n. 4, p. 639-652, 1977.
- TINBERGEN, J. *Income differences: recent research*. Oxford: North Holland Publishing, 1975.
- UZAWA, H. Optimal technical change in an aggregative model of economic growth. *International Economic Review*, v. 6, p. 18-31, 1965.

(Originais recebidos em julho de 2002. Revistos em outubro de 2002.)